Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **428,** 354–369 (2013) doi:10.1093/mnras/sts034

The high-redshift ($z > 3$) active galactic nucleus population in the 4-Ms *Chandra Deep Field*-South

F. Vito,[1,2]★ C. Vignali,[1,2] R. Gilli,[2] A. Comastri,[2] K. Iwasawa,[3] W. N. Brandt,[4,5] D. M. Alexander,[6] M. Brusa,[7] B. Lehmer,[8,9] F. E. Bauer,[10,11] D. P. Schneider,[4,5] Y. Q. Xue[12] and B. Luo[4,5]

[1]*Dipartimento di Astronomia, Università degli Studi di Bologna, Via Ranzani 1, 40127 Bologna, Italy*
[2]*INAF-Osservatorio Astronomico di Bologna, Via Ranzani 1, 40127 Bologna, Italy*
[3]*ICREA and Institut de Cincies del Cosmos (ICC), Universitat de Barcelona (IEEC-UB), Martí i Franquès 1, 08028 Barcelona, Spain*
[4]*Department of Astronomy and Astrophysics, The Pennsylvania State University, University Park, PA 16802, USA*
[5]*Institute for Gravitation and the Cosmos, The Pennsylvania State University, University Park, PA 16802, USA*
[6]*Department of Physics, University of Durham, South Road, Durham DHI 3LE*
[7]*Max Planck Institut für Extraterrestische Physik, Postfach 1312, 85741 Garching bei München, Germany*
[8]*The Johns Hopkins University, Homewood Campus, Baltimore, MD 21218, USA*
[9]*NASA Goddard Space Flight Center, Code 662, Greenbelt, MD 20771, USA*
[10]*Departamento de Astronomía y Astrofísica, Pontificia Universidad Católica de Chile, Casilla 306, Santiago 22, Chile*
[11]*Space Science Institute, 4750 Walnut Street, Suite 205, Boulder, CO 80301, USA*
[12]*Key Laboratory for Research in Galaxies and Cosmology, Department of Astronomy, University of Science and Technology of China, Chinese Academy of Sciences, Hefei, Anhui 230026, China*

Accepted 2012 September 18. Received 2012 September 15; in original form 2012 August 3

ABSTRACT

We present results from a spectral analysis of a sample of high-redshift ($z > 3$) X-ray-selected active galactic nucleus (AGN) in the 4-Ms *Chandra Deep Field*-South (CDF-S), the deepest X-ray survey to date. The sample is selected using the most recent spectroscopic and photometric information available in this field. It consists of 34 sources with median redshift $z = 3.7$, 80 median net counts in the 0.5–7 keV band and median rest-frame absorption-corrected luminosity $L_{2-10\,\mathrm{keV}} \approx 1.5 \times 10^{44}\,\mathrm{erg\,s^{-1}}$. Spectral analysis for the full sample is presented and the intrinsic column density distribution, corrected for observational biases using spectral simulations, is compared with the expectations of X-ray background (XRB) synthesis models. We find that ≈57 per cent of the sources are highly obscured ($N_\mathrm{H} > 10^{23}\,\mathrm{cm^{-2}}$). Source number counts in the 0.5−2 keV band down to flux $F_{0.5-2\,\mathrm{keV}} \approx 4 \times 10^{-17}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$ are also presented. Our results are consistent with a decline of the AGN space density at $z > 3$ and suggest that, at those redshifts, the AGN obscured fraction is in agreement with the expectations of XRB synthesis models.

Key words: methods: data analysis – techniques: imaging spectroscopy – surveys – galaxies: active – galaxies: high-redshift – X-rays: galaxies.

1 INTRODUCTION

The discovery of super massive black holes (SMBHs) at $z > 6$ with masses of the order of $10^9\,\mathrm{M_\odot}$ (e.g. Mortlock et al. 2011, the redshift record $z = 7.085$; Willott, McLure & Jarvis 2003; Fan et al. 2006a,b; Willott et al. 2009) implies that SMBHs were already in place when galaxies were only a few hundred Myr old. The formation of these massive objects and the physical conditions allowing a rapid and early growth are challenging problems. The origin of the SMBH seeds is also uncertain. Two classes of formation models are currently under discussion: light seeds, perhaps the remnants of Pop III stars ($\sim 10^2\,\mathrm{M_\odot}$; Madau & Rees 2001; Alvarez, Wise & Abel 2009; Volonteri & Begelman 2010, see also Whalen & Fryer 2012), possibly evolved as a 'quasi-star' (Begelman 2010), and heavy seeds, formed by monolithic collapse of massive gas clouds ($10^{4-5}\,\mathrm{M_\odot}$; e.g. Ball et al. 2011). Whatever the nature of the seed is, continuous Eddington-limited or even super-Eddington accretion on to these objects is required to reach the observed masses.

Such a high accretion rate requires a large amount of accreting material which, at the same time, may act as a screen against the emitted radiation (e.g. Hopkins et al. 2008). Supporting observational evidence was found for an increasing fraction of obscured active galactic nuclei (AGN) from $z = 0$ to $z \approx 2$–3 (La

★E-mail: fabio.vito@unibo.it

Franca et al. 2005; Ballantyne, Everett & Murray 2006; Treister & Urry 2006; Hasinger 2008; but see also Gilli et al. 2010, who dispute these results). At these redshifts, a large amount of gas is available in the host galaxy. The nuclear activity could be triggered by the gas falling towards the centre, when a 'wet' (i.e. gas-rich) merging event occurs. After an initial obscured ($N_{\rm H} > 10^{22}$ cm^{-2}) accretion phase, which in some models can be Compton-thick ($N_{\rm H} > 10^{24}$ cm^{-2}), the emitted radiation sweeps out the gas, revealing itself and acting as a negative feedback towards the star formation in the host galaxy (e.g. Brusa et al. 2005; Alexander et al. 2010; Feruglio et al. 2010; Sturm et al. 2011; Maiolino et al. 2012). The evolution of obscuration, in particular at high redshift, when most of the mass accretion was taking place, is a fundamental observable to probe this framework.

The AGN comoving space density in the optical and X-ray band is not constant over time: there are many more AGN at $z > 1$ than in the local Universe (e.g. Hasinger, Miyaji & Schmidt 2005). The evolution of the AGN population can be described by a phenomenological model known as luminosity-dependent density evolution model (LDDE model; e.g. Hasinger et al. 2005; Brandt & Alexander 2010): according to this model, the comoving space density of more luminous AGN ($L_{\rm X} > 10^{45}$ erg s^{-1}) peaks at $z \sim 2-3$ and then decreases exponentially to $z \sim 6$; for less luminous ones ($L_{\rm X} \lesssim 10^{45}$ erg s^{-1}) the peak arises at more recent times ($z \sim 1$–1.5) and then their density slowly decreases till the highest redshifts so far probed ($z \sim 3$). Therefore, more massive black holes formed and grew earlier than smaller ones. This sort of 'cosmic downsizing' is similar to that found for galaxies (e.g. Cowie, Songaila & Barger 1999; Thomas et al. 2005; Damen et al. 2009).

This scenario is, however, far from being complete. Particularly, the behaviour at higher redshifts ($z \geq 3$) is not constrained (especially at $L_{\rm X} < 10^{44}$ erg s^{-1}), implying severe uncertainties in AGN evolution during the early ages of the Universe. Furthermore, the space density and the role of heavily obscured AGN are largely unknown beyond the local Universe. In order to place constraints on the obscured AGN fraction and its evolution history, observables, such as the spectral properties and the number counts, are required to allow comparisons with evolutionary theoretical models.

Optical surveys are not suitable for these tasks since they are severely biased against obscuration. Deep X-ray surveys are required, because they are less affected by obscuration and can reach fluxes faint enough to sample the bulk of the high-redshift AGN population. Trichas et al. (2012) presented the largest (78 objects) sample of $z > 3$ X-ray selected AGN, using data from the Chandra Multi-wavelength Project (ChaMP) survey. A few studies have been performed on high-redshift AGN exploiting the deep surveys in the Cosmological Evolution Survey (COSMOS) field carried out with *XMM–Newton* (Brusa et al. 2009) and *Chandra* (Civano et al. 2011), and they were limited to relatively bright X-ray fluxes ($F_{0.5-2\,\rm keV} \gtrsim 10^{-15}$ and 3×10^{-16} erg cm^{-2} s^{-1}, respectively) and high 2–10 keV luminosities ($L_{2-10\,\rm keV} > 10^{44.2}$ and $10^{43.5}$ erg s^{-1}, respectively). Only recently, thanks to the 4-Ms *Chandra Deep Field*-South (CDF-S; Xue et al. 2011), the deepest X-ray survey to date, very faint fluxes (down to $F \approx 3.2 \times 10^{-17}/9.1 \times 10^{-18}/5.5 \times 10^{-17}$ erg cm^{-2} s^{-1} in the 0.5−8/0.5−2/2−8 keV band) could be reached. Early results on $z > 3$ population from the 4-Ms CDF-S data have been presented in Fiore et al. (2012). Lehmer et al. (2012) also presented the source counts of the $z > 3$ population as part of a more comprehensive work on $\log N - \log S$ of the X-ray population.

In this work, we analyse data from the 4-Ms CDF-S. In particular, we have selected a sample of 'bona-fide' high-redshift ($z > 3$) AGN using the most recent spectroscopic and photometric information available, and performed an X-ray spectral study to understand their properties and, possibly, constrain the early phases of AGN evolution. The spectral properties, column density distribution and number counts of the sample are derived, discussed and compared with previous works and model predictions.

Throughout this work we assume a $\Omega_{\rm m} = 0.3$, $\Omega_\Lambda = 0.7$ and $H_0 = 70\,{\rm km\,s^{-1}\,Mpc^{-1}}$ Universe (Spergel et al. 2003). The errors on the spectral parameters correspond to the 90 per cent confidence level (CL) for one parameter of interest (Avni 1976).

2 THE $z > 3$ AGN SAMPLE

We assembled a sample of X-ray-selected AGN at $z > 3$ in the 4-Ms CDF-S on the basis of both spectroscopic and photometric redshifts. Spectroscopic redshifts for $z > 3$ AGN were collected from Szokoly et al. (2004), Vanzella et al. (2008), Popesso et al. (2009), Silverman et al. (2010), Vanzella et al. (2010), Wuyts et al. (2009) and E. Vanzella (private communication[1]). A quality flag ('secure/insecure') is associated with them in Xue et al. (2011), except for spectroscopic redshifts from Vanzella et al. (2010), Wuyts et al. (2009) and E. Vanzella (private communication), which are not used by Xue et al. (2011) and for which we retrieved the original quality information. Photometric redshifts were gathered from Luo et al. (2010, which also provides secondary solutions), Cardamone et al. (2010), Rafferty et al. (2011), Santini et al. (2009), Wuyts et al. (2008), Taylor et al. (2009) and Dahlen et al. (2010).

We first cross-matched the X-ray positions of all sources in the 4-Ms CDF-S main catalogue (740 sources; Xue et al. 2011) with the positions of the sources in the catalogues which are not collected by Xue et al. (2011) using a 1 arcsec matching radius. Considering the other catalogues, we assumed the counterparts reported by Xue et al. (2011), which used a likelihood-ratio matching technique. We pre-selected a sample of high-redshift AGN candidates, including all sources which have a spectroscopic or photometric redshift $z > 3$ in at least one of the considered catalogues (96 sources). Given the large amount of available information, we built up a set of criteria to select the final high-redshift sample. A source is included in the sample if it fulfils one of the following requirements.

(1) A spectroscopic redshift flagged as 'secure' $z_{\rm spec,s} > 3$ is provided.

(2) A spectroscopic redshift flagged as 'insecure' $z_{\rm spec,i} > 3$ is available and it is upgraded to 'secure' by re-analysing the optical spectrum, checking the quality and reliability of the identified spectral features.

(3) $N_{\rm phot,\,>3} - N_{\rm phot,\,<3} \geq 3$, where $N_{\rm phot,\,>3}$ is the number of photometric redshifts larger than 3 and $N_{\rm phot,\,<3}$ is the number of photometric redshifts lower than 3, when no secure spectroscopic redshift and more than two photometric redshifts are available.

(4) When no secure spectroscopic redshift and two photometric redshifts are available, both photometric redshifts are larger than 3.

(5) When no secure spectroscopic redshift and only one photometric redshift are available, its 1σ lower bound (provided by the original catalogue) is larger than 3.

The selection procedure is shown in Fig. 1 as a flow chart. We will discuss in Section 4 the effects of relaxing the selection criteria on the sample size.

[1] These spectroscopic redshifts are based on a reanalysis of VIsible MultiObject Spectrograph (VIMOS) and FOcal Reducer and low dispersion Spectrograph (FORS2) spectra.

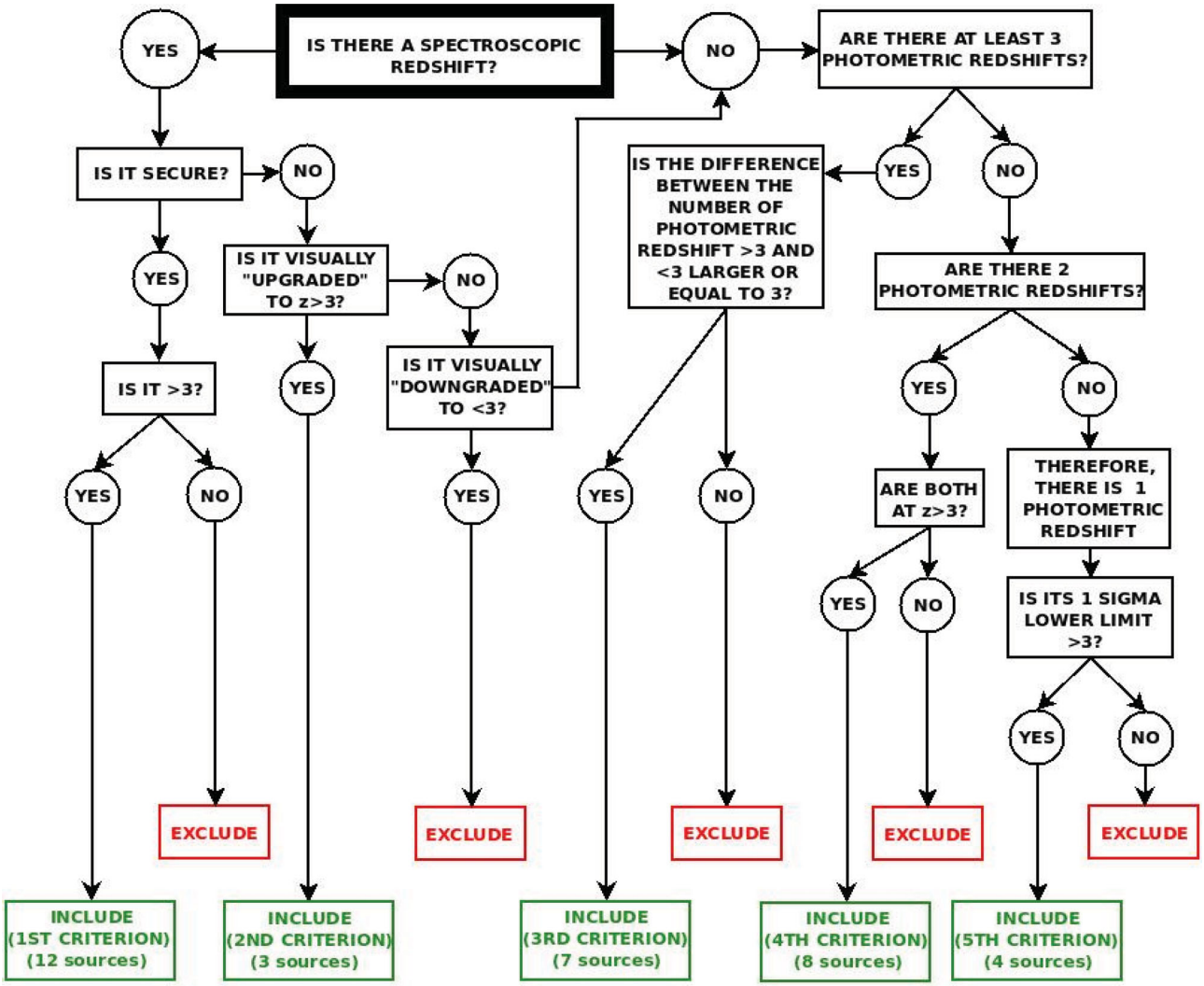

Figure 1. Selection criteria applied to the 96 pre-selected sources (which have at least one spectroscopic or photometric redshift $z > 3$) to be included in the high-redshift sample, visualized as a flow chart. The starting block has bold sides.

The final sample consists of 34 sources (see Table 1). A spectroscopic redshift is adopted for 15 sources (12 and three are included in the high-redshift sample on the basis of the first and second criteria, respectively) and a photometric redshift is adopted for 19 sources (7/8/4 sources are included in the sample on the basis of the third/fourth/fifth criterion, respectively). XID 392 has no redshift information in Xue et al. (2011) and a photometric redshift $z = 6.22$ from Santini et al. (2009), but no error on the redshift is provided. Therefore, we conservatively excluded this source from the sample. The sample presented in this work is a factor of $\sim$2 smaller than that discussed in Lehmer et al. (2012). This difference will be discussed in more details in Section 4.

When more than one redshift per source is available, we almost entirely followed the choice of Xue et al. (2011). Briefly, the adopted redshift in this work is in order of preference: secure (or upgraded to secure) spectroscopic redshift, photometric redshift from Luo et al. (2010) and photometric redshift from Cardamone et al. (2010). XID 331 has no redshift information in the catalogues collected by Xue et al. (2011), but the inclusion of the other above-mentioned catalogues allows that source to be included in the sample. We assumed photometric redshift from Dahlen et al. (2010) for it. No spectroscopic redshift is reported in Xue et al. (2011) for XID 235, 458, 528 and 386; redshift for these sources were obtained by E. Vanzella (XID 235 and 458, private communication), Vanzella et al. (2010, XID 528) and Wuyts et al. (2009, XID 386). A high-quality flag is associated with all these redshift in the original works; therefore, we consider them as 'secure'.

The redshift distribution of the sample is shown in Fig. 2. The mean and median redshifts are 4.0 and 3.7, respectively. Most of the sources with a spectroscopic redshift (13 out of 15) are at $z < 4$. We note the presence of three sources at $z > 5$, for which we investigated in more detail the redshift information. Two of them (XID 139 and 197) have only one photometric redshift ($z = 5.729$ and 6.071, respectively, from Luo et al. 2010, which also provides a secondary solution of $z = 4.385$ for XID 197), while XID 485 has two different photometric redshifts ($z = 7.62$ from Luo et al. 2010, with secondary solution $z = 3.309$, and $z = 4.42$ from Santini et al. 2009). These objects have no optical counterparts in the Great Observatories Origins Deep Survey (GOODS) (XID 139 and 485) or Galaxy Evolution from Morphologies and SEDs (GEMS) (XID 197, which is not in the GOODS field) observations, while they are significantly detected by Infrared Array Camera (IRAC) in the near-infrared (near-IR). Thus, a high-redshift ($z > 3$) solution appears probable.

Table 1. Main information and spectral parameters of the 34 high-redshift AGN.

XID (1)	RA[J2000] (2)	Dec.[J2000] (3)	z (4)	r (5)	C (6)	$N_{\rm H}$ (7)	CSTAT/d.o.f. (8)	$F_{0.5-2\,\rm keV}$ (9)	$F_{2-10\,\rm keV}$ (10)	$L_{2-10\,\rm keV}$ (11)	Counts (12)
27	52.96054	−27.87706	4.385	P8	5	68^{+41}_{-29}	71.7/113	2.45E−16	1.76E−15	2.80E44	77
100	53.01658	−27.74489	3.877	P8	3	38^{+27}_{-24}	69.0/92	1.48E−16	<7.55E−16	8.94E43	67
107	53.01975	−27.66267	3.808	P9	4	<71	51.6/62	4.80E−16	<1.40E−15	1.52E44	36
132	53.03071	−27.82836	3.528	P8	3	13^{+16}_{-11}	69.0/65	1.13E−16	<3.39E−16	3.12E43	57
139	53.03496	−27.67975	5.729	P8	5	96^{+33}_{-35}	92.8/101	5.71E−16	3.64E−15	1.02E45	111
150	53.03996	−27.79847	3.337	P9	3	45^{+49}_{-39}	17.3/31	<2.24E−17	1.58E−16	1.40E43	26
170	53.04746	−27.87047	3.999	P8	4	35^{+4}_{-4}	308/359	1.63E−15	7.49E−15	9.43E44	1472
197	53.05771	−27.93344	6.071	P8	5	86^{+31}_{-30}	88.2/124	6.24E−16	3.27E−15	1.02E45	134
235	53.07029	−27.84564	3.712	S7	1	<36	33.0/27	4.29E−17	7.95E−17	8.06E42	27
262	53.07854	−27.85992	3.66	S1	1	85^{+22}_{-20}	123.6/165	1.20E−16	1.53E−15	1.63E44	208
283	53.08467	−27.70811	3.204	P8	4	55^{+39}_{-22}	58.0/80	1.04E−16	9.54E−16	7.93E43	77
285	53.08558	−27.85822	4.253	P8	3	<6	29.7/31	5.71E−17	<9.13E−17	1.24E43	18
331	53.10271	−27.86061	3.78	P10	4	<23	48.1/36	4.61E−17	<7.39E−17	7.71E42	22
371	53.11158	−27.76789	3.101	P8	4	35^{+11}_{-10}	98.1/118	1.22E−16	7.70E−16	5.70E43	146
374	53.11204	−27.86072	3.724	P9	4	213^{+284}_{-114}	32.7/34	<3.59E−18	2.95E−16	4.14E43	22
386	53.11796	−27.73439	3.256	S6	1	<7	62.6/64	8.85E−17	<1.42E−16	1.06E43	40
403	53.12196	−27.93883	4.762	S2	1	185^{+148}_{-76}	66.9/73	<5.46E−17	1.30E−15	2.68E44	36
412	53.12442	−27.85169	3.7	S1	1	82^{+12}_{-12}	208.0/225	2.08E−16	2.46E−15	2.85E44	373
458	53.13854	−27.82128	3.474	S7	1	93^{+74}_{-70}	14.7/25	1.59E−17	2.52E−16	2.64E43	27
485	53.14658	−27.87103	7.620	P8	4	406^{+192}_{-114}	74.2/94	5.75E−17	8.86E−16	5.00E44	87
521	53.1585	−27.73372	3.417	P8	4	<23	47.6/50	8.91E−17	<2.40E−16	2.05E43	30
528	53.16158	−27.85606	3.951	S5	1	74^{+24}_{-24}	81.6/111	9.56E−17	8.93E−16	1.16E44	126
546	53.16533	−27.81419	3.064	S1	1	52^{+4}_{-4}	331.1/355	6.75E−16	6.26E−15	4.73E44	1072
556	53.17012	−27.92975	3.528	P8	3	97^{+9}_{-10}	285.6/293	7.59E−16	1.07E−14	1.14E45	718
563	53.17442	−27.86742	3.61	S1	1	6^{+2}_{-2}	257.6/307	2.07E−15	4.66E−15	4.45E44	1084
573	53.1785	−27.78411	3.193	S1	2	3^{+2}_{-2}	209.2/218	8.10E−16	1.65E−15	1.19E44	609
588	53.18467	−27.88103	3.471	S1	1	<2	115.5/153	6.45E−16	1.03E−15	8.41E43	281
642	53.20821	−27.74994	3.769	P9	3	<13	55.7/65	9.17E−17	<1.48E−16	1.54E43	18
645	53.20933	−27.88119	3.47	S3	2	15^{+2}_{-2}	371.4/384	3.16E−15	1.03E−14	9.24E44	2053
651	53.21529	−27.87033	4.658	P8	3	151^{+39}_{-35}	115.8/132	1.33E−16	2.38E−15	4.64E44	146
674	53.24004	−27.76361	3.082	S3	1	<7	120.3/146	6.95E−16	1.39E−15	8.56E43	199
700	53.2625	−27.86308	4.253	P8	5	18^{+16}_{-14}	73.8/90	7.43E−16	2.23E−15	3.11E44	102
717	53.28	−27.79892	4.635	S4	2	87^{+65}_{-51}	47.2/73	1.21E−16	1.02E−15	1.85E44	44
730	53.29587	−27.79317	3.724	S4	1	60^{+49}_{-39}	59.6/83	<1.81E−16	<1.46E−15	1.65E44	37

(1) Identification number from Xue et al. (2011); (2) right ascension and (3) declination from Xue et al. (2011); (4) adopted redshift (see Section 2); (5) reference for the adopted redshift – S1, S2, S3, S4, S5, S6 and S7: spectroscopic redshift from Szokoly et al. (2004), Vanzella et al. (2008), Popesso et al. (2009), Silverman et al. (2010), Vanzella et al. (2010), Wuyts et al. (2009) and E. Vanzella (private communication), respectively; P8, P9 and P10: photometric redshift from Luo et al. (2010), Cardamone et al. (2010) and Dahlen et al. (2010, weighted solution), respectively; (6) criterion satisfied for source inclusion in the high-redshift sample (see Section 2); (7) best-fitting $N_{\rm H}$ as derived in Section 3.2, in units of $10^{22}\,\rm cm^{-2}$. We define $N_{\rm H}$ to be constrained if the lower limit (at the 90 per cent CL) on the best-fitting value is larger than zero, otherwise we report its upper limit. (8) best-fitting value of Cash statistic over degrees of freedom; (9) soft-band (0.5–2 keV) and (10) hard-band (2–10 keV) flux or upper limits from the best-fitting spectral model, in units of $\rm erg\,cm^{-2}\,s^{-1}$; (11) intrinsic (i.e. absorption-corrected) rest-frame 2–10 keV luminosity, in units of $\rm erg\,s^{-1}$ (typical errors range from ∼10 per cent for the brightest sources to a factor of ∼2 for the faintest ones, fixing the redshift to the value reported in column 4); (12) full-band (0.5–7 keV) net (i.e. background-subtracted) counts, as obtained from the spectral extraction. The full-band net-count number is only representative, being the energy range considered during the spectral fitting narrower, in most cases, in order to maximize S/N.

3 DATA ANALYSIS

3.1 Spectral extraction procedure

Chandra data products for each of the 54 observations of the CDF-S are publicly available in the *Chandra* Data Archive.[2] We retrieved all data products (in particular, the evt2 files; i.e. the event list filtered on the good time intervals and status bits), but those related to the first two observations (for which the focal-plane temperature was −110°; Luo et al. 2008; Xue et al. 2011), and used them for spectral analysis. The sources were located in each of the 52 CDF-S images according to the coordinates provided by the main-source catalogue Xue et al. (2011).

We selected the extraction regions for the source and background using SAOIMAGE DS9 (Joye & Mandel 2003).[3] Despite the *Chandra* point spread function (PSF) distortion with the off-axis angle, we always chose circular regions, provided that most of the source counts

[2] http://cxc.harvard.edu/cda/

[3] http://hea-www.harvard.edu/RD/ds9/

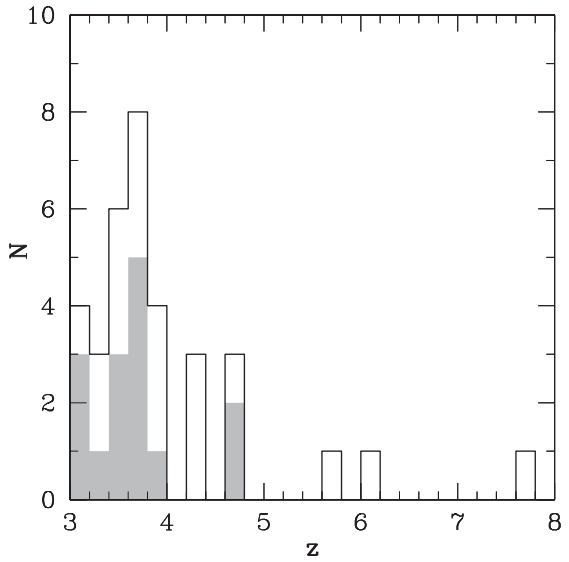

Figure 2. Redshift distribution of all 34 sources (solid line) and sources with spectroscopic redshift (shaded histogram) in the sample.

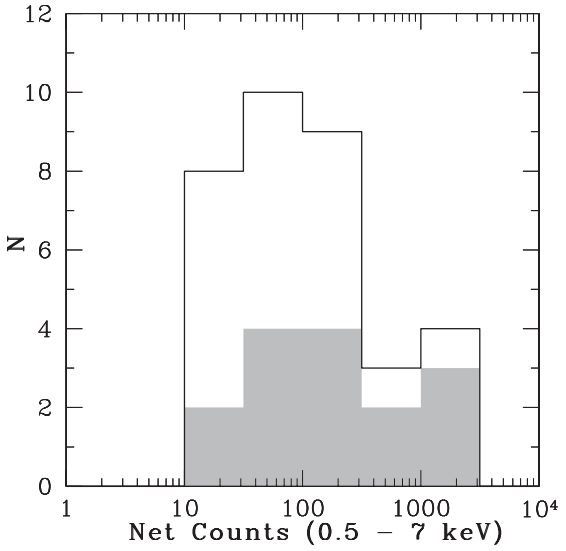

Figure 3. Full-band (0.5–7 keV) net counts distribution of all 34 sources (solid line) and sources with spectroscopic redshift (shaded histogram) in the sample.

are included in the selected regions [probably this choice slightly lowers the signal-to-noise ratio (S/N)]. Small radii are required for faint sources in order to include a low number of background counts and therefore to increase the S/N. Larger extraction radii can be used for bright sources. Also the position of the source in the field of view (f.o.v.) must be taken into account, because of the degradation of the PSF with the off-axis angle. For each source, we defined the same radius for the extraction region in all observations, considering the mean off-axis angle, provided by the main catalogue (Xue et al. 2011), and the relation between it and the encircled energy radius.

The off-axis angles for the sources in the sample are in the range of 1.42–10.1 arcmin. We chose the extraction radius for each source to obtain an acceptable trade-off between the number of counts and the S/N. As a general rule, for sources at off-axis angles $\lesssim$4 arcmin we defined the extraction radius corresponding to an encircled energy fraction (EEF) of 90 per cent at 1.49 keV. For sources at a large off-axis angle ($\gtrsim$7 arcmin), we considered radii in a range corresponding to 50–90 per cent EEF at 1.49 keV. For sources at intermediate off-axis angles (4–7 arcmin), we tried different extraction radii and chose those which provided the best trade-off between source and background counts. In a few cases we were forced to use smaller radii to avoid the contamination by sources close to the one we were considering. As a result, the extraction radii for the sample are in the range of 1.5–5 arcsec. Regarding the background, we chose extraction regions as close as possible to the source, preferably on the same chip, in areas where no other source was detected. In particular, we excluded all the regions around detected sources with radii a factor of $\geq$1.2 larger than the radii corresponding to 90 per cent EEF at 6.4 keV (which are a factor of $>$1.5 larger than the radii corresponding to 90 per cent EEF at 1.49 keV), in order to avoid contamination.

We used CIAO v4.3[4] to extract source and background spectra and to create the response files for each observation. In particular, we used the *specextract* tool which contains some improvements with respect to similar scripts in previous CIAO versions (e.g. it applies the energy-dependent aperture correction to the ARF file, accounting for the fraction of the PSF enclosed by a region). The spectra and response files of each source obtained for each observation were then added using tools provided in the FTOOLS package[5] (Blackburn 1995) (i.e. *mathpha*, *addarf* and *addrmf*), weighted by the exposure time of the individual observations.

The distribution of full-band (0.5–7 keV) net (i.e. background subtracted) counts (Fig. 3) is peaked near 100 counts (the mean and median values are 275 and 80, respectively). As expected, the bulk of the objects has a very low number of net counts: $\sim$74 per cent of sources in the sample have less than 200 net counts. In order to maximize the S/N, in most cases we fitted the spectra in a narrower energy range (0.5–5 keV), given that in the 5–7 keV band the background is high. All the spectra were fitted using the Cash statistic (Cash 1979) to estimate the best-fitting parameters.

3.2 Spectral analysis

The spectral fitting was performed with XSPEC v12.5.1n (Arnaud 1996).[6] We assumed an absorbed power law, parametrized by XSPEC models *powerlaw* and *zwabs*, as the starting spectral model. Galactic absorption was included assuming a column density in the range of (6–9) $\times$ 10^{19} cm^{-2} (Kalberla et al. 2005).

The mean photon index is $\langle\Gamma\rangle = 1.51$ with dispersion $\sigma = 0.28$ for the eight brightest sources ($>$250 net counts) and $\langle\Gamma\rangle = 1.63$ with $\sigma = 0.89$ for the whole sample. The average photon index is slightly flatter than what is typically measured in AGN samples ($\Gamma = 1.8$; e.g. Turner et al. 1997; Tozzi et al. 2006), but we emphasize that we are sampling the rest-frame 2$-$20 keV band, and the likely presence of a reflection hump could flatten the observed X-ray spectrum. However, the measured value is statistically consistent with $\Gamma = 1.8$. Reasonable accuracy was obtained for the best-fitting parameters only for the eight brightest sources (with a mean error on Γ and $N_{\rm H}$ of 14 and 45 per cent, respectively). The spectral parameters were not well constrained for almost all the other sources (error on Γ of 117 per cent; only upper limits on $N_{\rm H}$ for most objects) because of the poor spectral quality.

[4] http://cxc.harvard.edu/ciao/

[5] http://heasarc.nasa.gov/ftools/ftools_menu.html

[6] http://heasarc.nasa.gov/xanadu/xspec/

The difficulties in simultaneously constraining Γ and $N_{\rm H}$ for sources with few counts can be explained considering that these two spectral parameters have a significant degree of degeneracy. The problem is further complicated by the redshifted absorption cut-off, which moves outside the *Chandra* bandpass at $z \gtrsim 3$ even for significant degrees of obscuration, introducing further uncertainties.

To better constrain the column density, we fixed the photon index to $\Gamma = 1.8$ and repeated the spectral analysis on all the sources. This photon index is a widely used value for sources with low counting statistics, being considered the typical slope for AGN power-law emission in X-rays (e.g. Turner et al. 1997; Tozzi et al. 2006). The best-fitting parameters, fluxes and intrinsic luminosities derived using this spectral model are reported in Table 1. Similar results are obtained fitting simultaneously the source and background spectra. All the X-ray spectra of the sources, fitted using this model, are shown in Appendix B. The effects of assuming a flatter photon index ($\Gamma = 1.6$) will be discussed in Sections 3.3 and 4.

An anticorrelation between the column density and the AGN intrinsic luminosity is reported in several works (e.g. Steffen et al. 2003; Ueda et al. 2003; Hasinger 2004; La Franca et al. 2005; Hasinger 2008; Brusa et al. 2010), possibly linked to radiative and mechanical feedback, sweeping away the obscuring gas and dust. We do not obtain a significant correlation between these two observables (see Fig. 4). However, the luminosity range ($43 \leq \log L_{2-10\,\rm keV} \leq 45$) is likely too narrow and the sample size too small to find such relation.

Fig. 5 reports the best-fitting $N_{\rm H}$ plotted against the adopted redshift for each source. We note that the majority ($\approx$68 per cent) of the sources are highly obscured ($N_{\rm H} > 10^{23}\,{\rm cm}^{-2}$). Furthermore, there is a hint for an increasing trend of column density with redshift, as found in other works (e.g. Treister & Urry 2006). However, a possible bias due to the high redshift and low counting statistics can be present and will be investigated in the next section. The trend could be driven by the three sources at $z > 5$, which have the most uncertain redshift information (see Section 2). If they were placed at $z = 3.7$ (the median redshift of the sample), the best-fitting $N_{\rm H}$ would be $50^{+23}_{-18} \times 10^{22}$, $42^{+20}_{-24} \times 10^{22}$ and $118^{+41}_{-28} \times 10^{22}\,{\rm cm}^{-2}$ for XID 139, 197 and 485, respectively (i.e. they would still be in the highly obscured class). Assuming the photometric redshift ($z =$ 4.42) from Santini et al. (2009) for XID 485, the only source in the sample at $z > 5$ for which more than one photometric redshift is provided by the catalogues that we considered, the best-fitting column density would be $145^{+52}_{-35} \times 10^{22}\,{\rm cm}^{-2}$ (Fig. 5, open circle). We recall that the low photon-counting statistics allow us to assume only a simple spectral model, which does not take into account complex absorption structures (e.g. Bianchi, Maiolino & Risaliti 2012), which could affect the results on the column density.

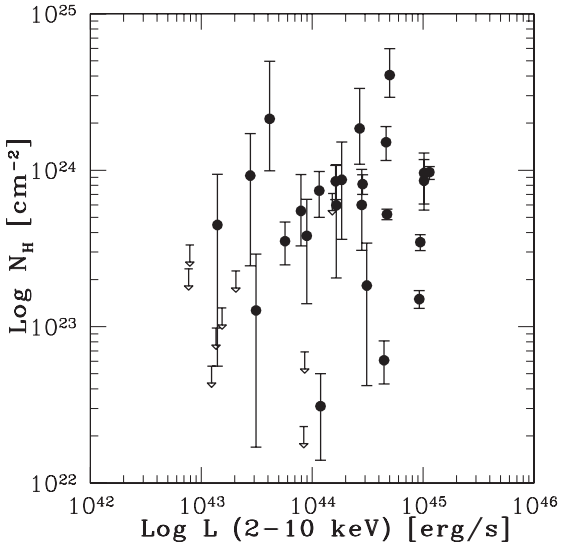

Figure 4. Best-fitting $N_{\rm H}$ as a function of the hard-band intrinsic luminosity for all the sources in the sample. Filled circles represent constrained values, while upper limits are reported as downward-pointing arrows.

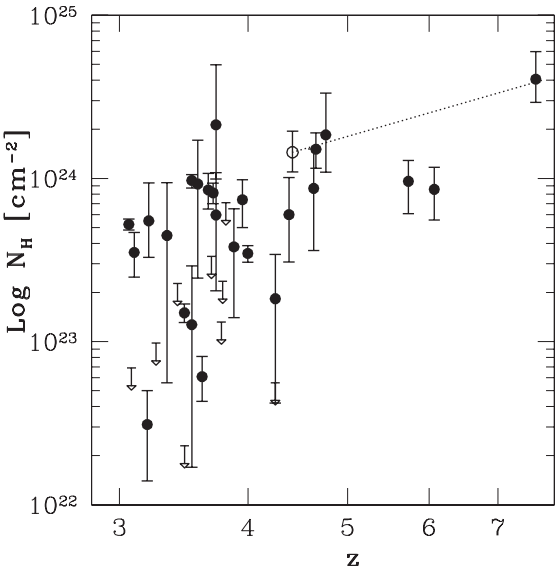

Figure 5. Best-fitting $N_{\rm H}$ as a function of redshift for all the sources in the sample. Filled circles represent constrained values, while upper limits are reported as downward-pointing arrows. The dashed line connects the best-fitting $N_{\rm H}$ of XID 485 at $z = 7.62$, as reported in Table 1, with the best-fitting value which would be derived if it were placed at $z = 4.42$ (open circle). See text for details.

The four sources with $N_{\rm H} > 10^{24}\,{\rm cm}^{-2}$ were also fitted with the XSPEC model *plcabs* (power-law emission transmitted through a spherical distribution of cold and dense matter, taking into account Compton scattering; Yaqoob 1997), which reproduces X-ray emission from Compton-thick sources better than a simple absorbed power law, but no significant change in the best-fitting spectral parameters was found.

We also investigated the possible presence of a neutral iron Kα line in the spectra by adding a narrow ($\sigma = 0.01$ keV) Gaussian component (*zgauss* model in XSPEC) at 6.4 keV rest frame. We could only find upper limits to the rest-frame equivalent width (EW) for most of the sources, of the order of a few 100 eV on average. This result is easily explained given that the observed EW scales with $(1 + z)^{-1}$ and therefore only loose constraints can be placed to the line in spectra of high-redshift AGN (especially if the S/N and the number of net counts are low and the accuracy in the redshift measurement is limited). The only source for which the iron line was easily visible in the spectrum and statistically significant is XID 412 (i.e. CDF-S 202 in Norman et al. 2002). In this case, we allowed the rest-frame energy of the line to vary during the spectral fitting in a relatively large range. The best-fitting rest-frame energy, EW and $N_{\rm H}$ that we derived for this source ($E =$ 6.53 keV, EW $= 738^{+465}_{-451}$ eV and $N_{\rm H} = 91^{+14}_{-13} \times 10^{22}\,{\rm cm}^{-2}$) are in good agreement with those found by previous *Chandra* (Norman et al. 2002) and *XMM–Newton* (Comastri et al. 2011) observations. The soft ($F_{0.5-2\,\rm keV} = 1.76 \times 10^{-16}\,{\rm erg\,cm}^{-2}\,{\rm s}^{-1}$) and hard ($F_{2-10\,\rm keV} = 2.41 \times 10^{-15}\,{\rm erg\,cm}^{-2}\,{\rm s}^{-1}$) fluxes are consistent with

those reported by Tozzi et al. (2006) using the 1-Ms CDF-S data, although they assumed a different spectral model ($\Gamma = 1.8$ and $N_{\rm H} = 1.5 \times 10^{24}$ cm^{-2}).

3.3 $N_{\rm H}$ distribution

The distribution of the best-fitting $N_{\rm H}$ (Fig. 6) derived from spectral analysis, assuming an absorbed power law with fixed $\Gamma = 1.8$ as fitting spectral model, strongly suggests the presence of an important fraction (≈68 per cent) of highly obscured sources ($N_{\rm H} > 10^{23}$ cm^{-2}). This result could be affected by a systematic $N_{\rm H}$ overestimate due to the high redshift and low counting statistics, characteristic of the sample: the photoelectric cut-off in spectra of mildly obscured ($10^{22} < N_{\rm H} < 10^{23}$ cm^{-2}) sources is redshifted out of the *Chandra* bandpass already at $z = 3$–4, making very difficult the search for the best-fitting $N_{\rm H}$. Indeed, this type of spectrum might be indistinguishable from a spectrum of an unabsorbed source at high redshift. Furthermore, statistical fluctuations, enhanced by the low fluxes, are expected in the spectra and could simulate a spurious photoelectric cut-off at high rest-frame energies (Tozzi et al. 2006).

In order to estimate the probability (P_{ij}) that the X-ray spectrum of a source with intrinsic $N_{\rm H}$ in a given range j is fitted with a best-fitting column density constrained in a range i, we made extensive use of simulations. As in Table 1, we define $N_{\rm H}$ to be constrained if the 90 per cent CL lower limit on the best-fitting value is larger than zero. We divided the sample in five $N_{\rm H}$ bins, as reported in Table2. The number of sources N_i observed in each absorption bin i is then given by

$$N_i = \sum_j x_j P_{ij}, \quad (1)$$

where x_j is the number of sources with intrinsic $N_{\rm H}$ in bin j and N_i is the number of sources observed with best-fitting $N_{\rm H}$ in bin i, for $i, j =$ A, B, C, D and E. In this framework, N_i is known (black histogram in Fig. 6) and P_{ij} can be derived from simulations.

We simulated 1000 spectra with $\Gamma = 1.8$ and 100 net counts (close to the median value of our sample), intrinsically obscured by a column density in the jth $N_{\rm H}$ bin (for each bin we used the value reported in the third column of Table 2), considering sources

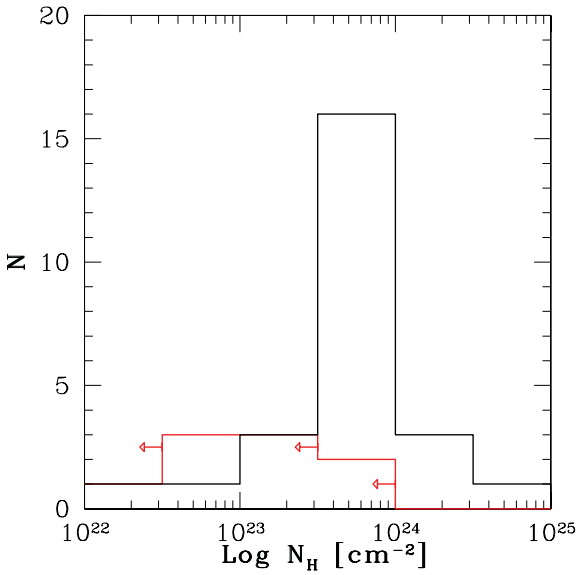

Figure 6. Observed distribution of the best-fitting $N_{\rm H}$. The black line represents the constrained values; the red line with leftward pointing arrows represents the upper limits.

Table 2. $N_{\rm H}$ bins used in simulations.

i,j	Range: $\log(\frac{N_{\rm H}}{\rm cm^{-2}})$	Sim: $\log(\frac{N_{\rm H}}{\rm cm^{-2}})$
A	<21	0.0
B	21–22	21.5
C	22–23	22.5
D	23–24	23.5
E	>24	24.5

Sources in the ith or jth $N_{\rm H}$ bin (first column), which accounts for the $N_{\rm H}$ range shown in the second column, are simulated assuming the column density reported in the third column.

Table 3. Final probability factors P_{ij} that a source with intrinsic $N_{\rm H}$ in a given bin j is observed in a bin i. These are derived from simulations using $\Gamma = 1.8$ and 1.6.

P_{ij}	$j =$ (ABC)	$j =$ D	$j =$ E
	$\Gamma = 1.8$		
$i =$ (ABC)	0.361	0.0	0.0
$i =$ D	0.639	1.0	0.0
$i =$ E	0.0	0.0	1.0
	$\Gamma = 1.6$		
$i =$ (ABC)	0.147	0.0	0.0
$i =$ D	0.853	1.0	0.0
$i =$ E	0.0	0.0	1.0

at $z = 4$ (similar to the characteristic redshift of our sample), and simply counted how many times the simulated spectra are fitted with a best-fitting $N_{\rm H}$ constrained in the ith bin (see Appendix A for the detailed description of the simulation procedure). Bins A, B and C are indistinguishable at $z > 3$ and hence are merged into a single bin (representing unobscured or mildly obscured sources with $N_{\rm H} < 10^{23}$ cm^{-2}), hereafter referred to as (ABC).

The value of each P_{ij} is shown in Table 3. We immediately note that no spurious effect seems to be important in bins D and E (i.e. spectra of intrinsically strongly obscured sources are efficiently fitted by a model obscured by a column density within the correct $N_{\rm H}$ bin) while, as expected, the best-fitting $N_{\rm H}$ is overestimated in most cases (≈64 per cent) when the intrinsic value is in the bin (ABC).

By inverting equation (1), the intrinsic $N_{\rm H}$ distribution can now be derived (Fig. 7): $x_{\rm (ABC)} \approx 5.5$, $x_{\rm D} \approx 15.5$ and $x_{\rm E} = 4$, to be compared with the observed distribution: $N_{\rm (ABC)} = 2$, $N_{\rm D} = 19$ and $N_{\rm E} = 4$. If we conservatively count all the nine sources for which only upper limits could be obtained (red line in Fig. 6) in the first bin, the resulting $N_{\rm H}$ distribution is the one shown in Fig. 8. We added the predictions of the Gilli, Comastri & Hasinger (2007, hereafter GCH07) X-ray background (XRB) synthesis model computed with the POMPA[7] tool for each $N_{\rm H}$ bin, considering sources with luminosity $42 \leq \log\frac{L_{0.5-2\,\rm keV}}{\rm erg\,s^{-1}} \leq 47$ in a redshift range between $z_{\rm min} = 3$ and $z_{\rm max} = 8$. A high-redshift exponential decline in the space density of AGN, similar to that of bright quasars (Gilli et al. 2011), is also considered. The assumption of the decline affects only the normalization and not the shape of the distribution. The predictions are corrected for the sky coverage (i.e. the sky area in physical units covered by the survey at different fluxes, see next section), to take into account the smaller sky area covered by the survey at low fluxes. Obscured sources have a flux lower than unobscured or less

[7] POrtable Multi-Purpose Application for AGN counts (http://www.bo.astro.it/∼gilli/counts.html).

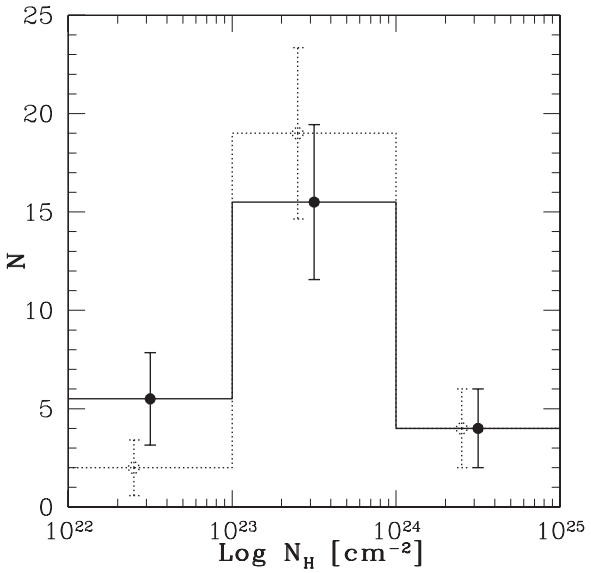

Figure 7. Intrinsic N_H distribution (solid line) with statistical errors, derived correcting the observed one (dotted line, plotted in the same bins, with statistical errors slightly shifted for visual purpose), resulting from spectral analysis assuming $\Gamma = 1.8$, with the probability factors obtained from simulations (see Section 3.3 and Appendix A). Only constrained values of N_H are considered.

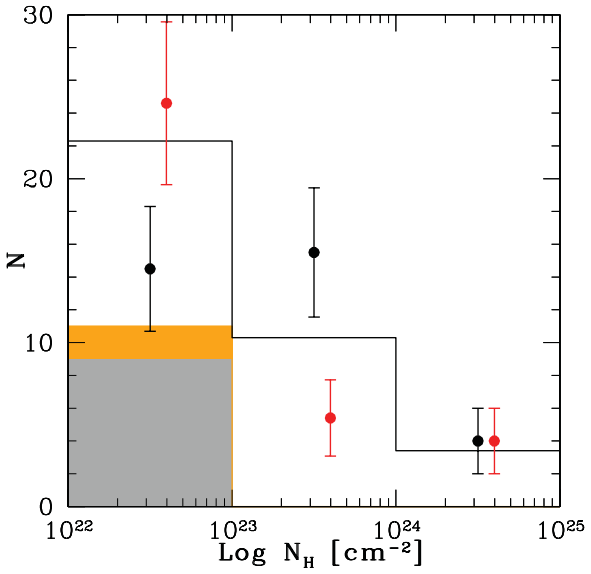

Figure 8. Intrinsic N_H distribution assuming $\Gamma = 1.8$ (black points) and $\Gamma = 1.6$ (red points, which are slightly shifted from the centre of the bins for visual purposes). The first bin includes sources with $N_H < 10^{23}$ cm^{-2} (i.e. consistent with no or 'moderate' absorption) and all the sources for which only upper limits on N_H were obtained (nine and 11 sources in the two cases, represented as grey and orange shaded areas, respectively). The third bin contains sources with $N_H > 10^{24}$ cm^{-2}. Poissonian errors are computed for the number of sources in each bin. The histogram is the predictions computed with the POMPA tool and corrected for the sky coverage of the survey (see text). The predictions are not normalized to the observed number of sources.

obscured ones with the same luminosity at the same redshift and, hence, can be detected over a smaller, deeper area.

The N_H distribution, corrected for the spurious overestimate of the column density, shows a number of strongly obscured sources that is slightly larger than the predictions of the model we considered. Conversely, the number of unobscured or mildly obscured sources is smaller than the predicted number. However, we find an agreement between the observations and the predictions at the 2σ CL (the χ^2 test returns a value of 6.0 for three degrees of freedom and no free parameter).

Following the results of Section 3.2 where, considering an absorbed power law with Γ free to vary as starting spectral model, we derived a mean photon index $\langle\Gamma\rangle \approx 1.6$, we investigated the effects of assuming a flatter photon index on the column density distribution. Therefore, we fitted again the X-ray spectra of all sources in the sample using the same spectral model with photon index fixed to $\Gamma = 1.6$. The resulting N_H distribution was corrected for spurious overestimates using the probability factors (Table 3) obtained from simulations, similar to the $\Gamma = 1.8$ case, but now using $\Gamma = 1.6$ to simulate and fit the spectra. Red points in Fig. 8 represent the intrinsic N_H distribution in the $\Gamma = 1.6$ case. We derived upper limits on N_H for two sources (XID 132 and 700, which were previously counted in bin D), while their best-fitting column density was constrained using $\Gamma = 1.8$. This result produces a different number of sources in the two shaded histograms in Fig. 8. Therefore, the number of objects in each bin of the corrected N_H distribution is, in this case, $x_{(ABC)} \approx 13.6$, $x_D \approx 5.4$ and $x_E = 4$. If we place all the 11 sources with an upper limit on N_H in the first bin, as in Fig. 8, the distribution is fully consistent within 2σ with the prediction of the model ($\chi^2 = 4.75$ with three degrees of freedom and no free parameter). We underline that the model predictions are computed with the same level of incompleteness as the observed sample.

3.4 The log N – log S of the high-redshift sample

We derived the log N–log S of the high-redshift sample in the soft band, where the survey sensitivity is the highest, folding the number N of sources with flux larger than S with the survey sky coverage at that flux. We assumed the soft fluxes reported in Table 1 derived from our spectral analysis. The sky coverage was obtained by multiplying the nominal sky area covered by the survey ($\approx$0.13 deg^2; Xue et al. 2011) and the fraction of the f.o.v. covered at different fluxes, derived applying the CIAO tool DMIMGHIST on the sensitivity map (Xue et al. 2011), which assigns to each point of the f.o.v. the flux limit of the survey at that position. In order to be consistent with the computation of the sky coverage (derived in the 0.5–2 keV band), we excluded four ($\approx$12 per cent of the total sample) sources which were not detected in the soft band (see Table 1).

In Fig. 9 we compare our data with those provided by Fiore et al. (2012) and Lehmer et al. (2012) for the 4-Ms CDF-S and with the *XMM* Brusa et al. (2009) and *Chandra* Civano et al. (2011) COSMOS data. The predictions of the GCH07 XRB synthesis model (solid and dashed lines) and the Aird et al. (2010) X-ray luminosity function (XLF) evolutionary model (dotted line) are also shown. The GCH07 model is based on an LDDE model of the XLF with redshift, and it allows the presence of a high-redshift ($z > 2.7$) decline in the AGN space density to be taken into account. The Aird et al. (2010) model assumes a luminosity and density evolution (LADE) of the XLF.

Our results for the $z > 3$ sample (top panel) are in good agreement with the GCH07 model predictions assuming an exponential decline in high-redshift AGN space density (continuous line), down to flux $F_{0.5-2\,keV} \approx 4 \times 10^{-17}$ erg cm^{-2} s^{-1}, a factor of $\approx$10 fainter with respect to previous surveys (i.e. down to $L_X \approx 10^{43}$ erg s^{-1}, the minimum luminosity detectable at $z > 3$ in the 4-Ms CDF-S): we extend to lower fluxes the results obtained in the COSMOS field by Brusa et al. (2009) and Civano et al. (2011), confirming an evolution

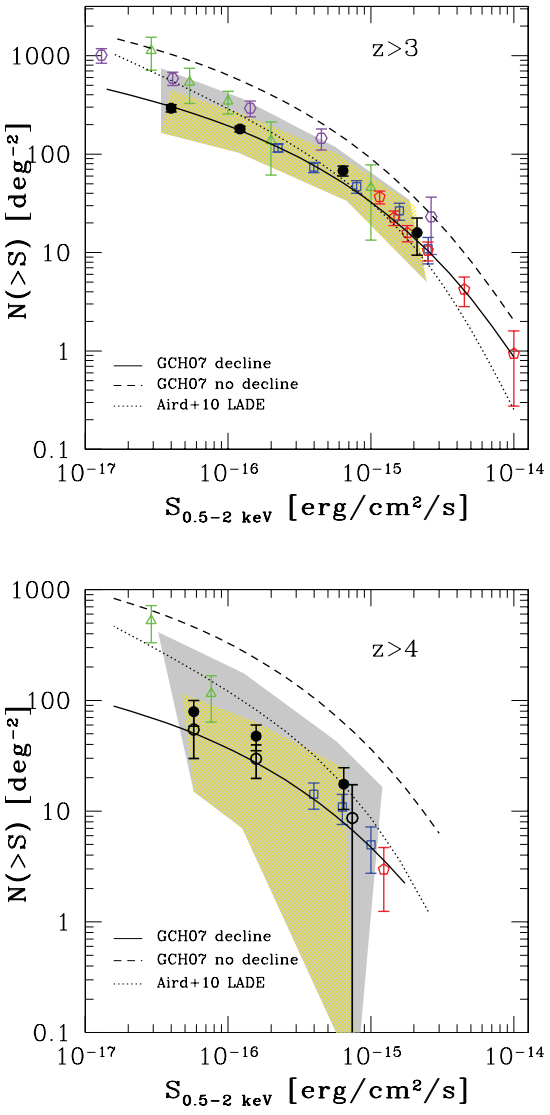

Figure 9. Binned $\log N - \log S$ of the high-redshift sample, considering all sources ($z > 3$, top panel) and only those at $z > 4$ (bottom panel), with associated Poissonian errors. Black filled circles are from this work, green triangles are from Fiore et al. (2012), blue squares are from Civano et al. (2011), red pentagons are from Brusa et al. (2009) and purple hexagons are from Lehmer et al. (2012). Empty black circles in the bottom panel represent our data if the three sources at the highest redshifts were placed at $3 < z < 4$ (see text). The predictions of the GCH07 XRB synthesis model, computed with the POMPA tool, either assuming (solid line) or not (dashed line) an exponential decline in the AGN space density at high redshift, and Aird et al. (2010) model are also shown. The yellow shaded area represents the regions in which the $\log N - \log S$ would be placed if more conservative or shallower selection criteria were used as well as the grey area, which also includes the 39 sources with no redshift in the considered catalogues (see Section 4).

at high redshift similar to that of the more luminous AGN (Gilli et al. 2011).

Considering only the $z > 4$ subsample (Fig. 9, bottom panel, filled circles), our data lie above the model predictions. We include here the presence of three sources at $5 < z \lesssim 7.7$, whose redshifts are determined on the basis of relatively uncertain photometric information. If these sources were placed at $3 < z < 4$, a good agreement would be obtained (Fig. 9, bottom panel, open circles).

The scenario without the decline (Fig. 9, dashed lines) is excluded by our data both at $z > 3$ and $z > 4$.

4 DISCUSSION

The $N_{\rm H}$ distribution derived from spectral analysis, corrected for the overestimate of the column density, suggests the presence of a significant fraction (∼57 per cent) of strongly obscured ($N_{\rm H} > 10^{23}\,{\rm cm}^{-2}$) AGN at $z > 3$. This result was obtained by fitting the X-ray spectra with an absorbed power law with $\Gamma = 1.8$, i.e. the typical photon index observed in large AGN samples (e.g. Turner et al. 1997; Tozzi et al. 2006). Adopting $\Gamma = 1.6$, we derived a slightly different intrinsic $N_{\rm H}$ distribution, with more unobscured sources, with respect to the previous case. Since the current data are sampling a rest-frame energy range up to at least $E_{\rm int} \approx 20$ keV, a hardening of the X-ray spectra due to reflection might be expected at observed energies $E_{\rm obs} \approx 4$–5 keV. However, the low photon-counting statistics prevented us from using relatively complex spectral models and the $\Gamma = 1.6$ index may simply incorporate the hardening due to reflection. Therefore, the 'true' $N_{\rm H}$ distribution is expected to lie between the two distributions derived assuming $\Gamma = 1.8$ and 1.6 (Fig. 8). To further investigate the possible impact of a reflection component, we computed the stacked, rest-frame, X-ray spectra (using the method described in Iwasawa et al. 2012) dividing the sample in four $N_{\rm H}$ bins ($\frac{N_{\rm H}}{10^{23}\,{\rm cm}^{-2}} < 1.3$; $1.3 < \frac{N_{\rm H}}{10^{23}\,{\rm cm}^{-2}} < 6$; $6 < \frac{N_{\rm H}}{10^{23}\,{\rm cm}^{-2}} < 9$; $\frac{N_{\rm H}}{10^{23}\,{\rm cm}^{-2}} > 9$; see Table 1). The resulting best-fitting slope, obtained by a joint fit of the various $N_{\rm H}$ bins and representing the intrinsic continuum above 10 keV, is $\Gamma \simeq 1.69^{+0.07}_{-0.07}$. This slope is suggestive of a limited contribution from reflection and is well within the values adopted to compute the absorption distribution (Fig. 8). Independently from the assumed spectrum, the data are in agreement at the 2σ CL with the prediction of the GCH07 XRB synthesis model, assuming an exponential decline in the AGN space density at $z > 2.7$ (Fig. 8). The best-fitting column densities of the 19 high-luminosity ($L_{2-10\,{\rm keV}} > 10^{44}\,{\rm erg\,s}^{-1}$) sources are well constrained at high values (∼84 per cent of them has nominal $N_{\rm H} > 10^{23}\,{\rm cm}^{-2}$). Applying the correction factors derived assuming $\Gamma = 1.8$ (see Section 3.3), we derived that the intrinsic fraction of obscured sources ($N_{\rm H} > 10^{23}\,{\rm cm}^{-2}$) in this subsample is ∼66 per cent.

The GCH07 model assumes that the fraction of obscured AGN does not evolve with redshift. A possible evolutionary scenario emerging from recent works suggests that this behaviour could be valid only for quasars (QSOs) with $L_{2-10\,{\rm keV}} \leq 10^{44}\,{\rm erg\,s}^{-1}$ (see Gilli et al. 2010), which is the median luminosity of our sample. Indeed, evidences are found that the activity of luminous ($L_{2-10\,{\rm keV}} \geq 10^{44}\,{\rm erg\,s}^{-1}$) AGN, triggered by major gas-rich mergers, occurs in short bursts (with time-scales of the order of ∼0.01 Gyr; e.g. Alexander et al. 2005). In this case, the gas accretion is chaotic (Hopkins et al. 2008) and may produce a large covering factor and column densities. Since the merging rate increases with redshift, the fraction of obscured AGN may then increase as well, as reported in e.g. La Franca et al. (2005), Hasinger (2008) and Iwasawa et al. (2012). At lower luminosities, the accretion on to the SMBH is probably driven by secular processes (∼1 Gyr; e.g. Daddi et al. 2007; Elbaz et al. 2011) and the accretion is expected to be smoother and symmetrical. Therefore, the fraction of obscured AGN would be quite constant with redshift, given that obscuration would be a function of the system geometry only. Our results fit well in this framework.

Fig. 9 shows the number counts derived in this work compared to previous determinations from the literature and theoretical models. The CDF-S source counts down to $F_{0.5-2\,{\rm keV}} \approx 4 \times 10^{-17}\,{\rm erg\,cm}^{-2}\,{\rm s}^{-1}$ lie on the XRB model if an exponential decline of the AGN XLF is assumed, similar to *XMM* (Brusa et al. 2009) and *Chandra* (Civano et al. 2011) COSMOS results at brighter fluxes. In particular, we note the agreement at bright

fluxes with the results from COSMOS surveys, for which a major effort was made to obtain spectroscopic and photometric redshifts (Brusa et al. 2009; Salvato et al. 2009, 2011, and references therein). Our results are consistent also with the Aird et al. (2010) model if the least conservative bounds (grey area in Fig. 9) are assumed (see below for the computation of the error budget). At fluxes $2 \times 10^{-16} \leq F_{0.5-2\,\mathrm{keV}} \leq 2 \times 10^{-15}$ erg cm^{-2} s^{-1} (i.e. for the brightest sources in the high-redshift sample), the prediction of this model is very similar to the GCH07 model, but at fainter fluxes the two models deviate significantly. Our results better trace the GCH07 model, while data from Fiore et al. (2012) and Lehmer et al. (2012) are placed on the Aird et al. (2010) model.

There is consistency in the number counts that we present with data from Fiore et al. (2012) and Lehmer et al. (2012) down to $F_{0.5-2\,\mathrm{keV}} \approx 10^{-16}$ erg cm^{-2} s^{-1}, while our results are a factor of $\sim$2–3 lower at fainter fluxes. This is not surprising considering that different selection techniques are applied, ours being more conservative than the ones used in those works. Indeed, the $\log N - \log S$ in Fiore et al. (2012) is derived for 54 optically selected sources at $z > 3$, using spectroscopic and photometric redshifts and dropout techniques and applying a new detection algorithm on the 4-Ms CDF-S observations, which made possible to reach a flux limit of $F_{0.5-2\,\mathrm{keV}} \approx 1.2 \times 10^{-17}$ erg cm^{-2} s^{-1}. The samples considered in Lehmer et al. (2012) and in this work are collected from the same catalogue of X-ray-selected sources. However, Lehmer et al. (2012) assumed the redshift adopted in Xue et al. (2011), while we collected additional information from other catalogues and used a stricter selection procedure (as described in Section 2). Therefore, we rejected a number of sources which are included in Lehmer et al. (2012), and we were able to collect redshift information for some objects which have no redshift in Xue et al. (2011), as in the case of XID 331. We would obtain results largely consistent with Lehmer et al. (2012) using their sample.

To compute the 'error budget' on the number counts, we investigate how many sources would be included in the sample if we relax the third, fourth and fifth criteria described in Section 2. Relaxing the third criterion and including in the final sample soft X-ray sources for which $N_{\mathrm{phot},>3} - N_{\mathrm{phot},<3} \geq 0$, the net gain in the final sample would be of four sources. As for the fourth criterion, 14 sources were excluded from the sample because only one of the two available photometric redshift is at $z > 3$. We consider these objects to have a 50 per cent chance of being at $z > 3$, and randomly include seven of them in the error budget. Finally, if we relax the fifth criterion and include in the sample all sources with only one photometric redshift, with the condition of being at $z > 3$, regardless of its error, we would add four sources to the final sample. If all these changes were adopted, the sample would consist of 49 sources.

The gold shaded area in Fig. 9 indicates the regions in which the $\log N - \log S$ would lie considering the error budget.[8] For the $z > 3$ number counts (upper panel), the lower bound is obtained if only sources selected via the first, second and third criteria of Section 2 are included in the high-redshift sample (i.e. sources with spectroscopic redshift or $N_{\mathrm{phot},>3} - N_{\mathrm{phot},<3} \geq 3$). This is clearly a very conservative assumption. The upper bound is obtained if we relax the third, fourth and fifth criteria, as discussed above, adding 15 sources in total.

Four out of these 15 sources would be at $z > 4$: two would be selected relaxing the fifth criterion and two relaxing the fourth criterion. As previously described, we randomly assumed half of the latter sources to be at $z > 4$. The resulting three sources were added to the $z > 4$ sample to compute the upper bound of the gold shaded area (bottom panel in Fig. 9), which accounts for the error budget of the $z > 4$ number counts. The lower bound is obtained, as before, considering only the first, second and third criteria, and consists of one source with a spectroscopic redshift $z > 4$ (XID 717, while we discarded XID 403 because it is not detected in the soft band) and two sources selected through the third criterion. As expected, because of the low number of sources detectable at $z > 4$, only loose constraints can be placed on the $\log N - \log S$. Larger samples of X-ray-selected AGN at $z > 4$ would be needed to draw more solid conclusions.

39 sources detected in the soft band in Xue et al. (2011) have no redshift in any catalogue that we considered. A not negligible fraction of them are probably AGN at $z > 3$. The grey shaded regions in Fig. 9 show how the gold area would enlarge if all these sources would be accounted for. Therefore, these areas represent the hypothetical and unlikely cases in which all the 39 sources were at $z > 3$ ($z > 4$).

5 CONCLUSIONS

The detection of a significant number of $z > 3$ AGN down to $L_{\mathrm{X}} \approx 10^{43}$ erg s^{-1} is made possible by the faint flux limit reached in the 4-Ms CDF-S observations (Xue et al. 2011). Moreover, X-rays are less biased against obscuration with respect to the optical band. The large multiwavelength coverage of the CDF-S allows us to retrieve spectroscopic and photometric redshifts for most of the sources in the field. In this framework, physical and evolutionary properties of a relatively large sample of high-redshift AGN can be studied.

The sample consists of 34 sources in the range of $3 < z \leq 7.6$, with a median redshift of $z = 3.7$. About 45 per cent of them (15 sources) have a spectroscopic redshift.

X-ray spectra of all the sources were extracted from the 4-Ms CDF-S observations and fitted with an absorbed power-law spectral model with photon index fixed to $\Gamma = 1.8$. The low photon-counting statistics prevents us from assuming a more complex spectral model: the median full-band (0.5–7 keV) net-count number is 80. The median rest-frame absorption-corrected luminosity of the sample is $L_{2-10\,\mathrm{keV}} \approx 1.5 \times 10^{44}$ erg s^{-1}.

We summarize the conclusions reached in this work as follows.

(i) The observed column density distribution is strongly peaked between $23 < \log(\frac{N_{\mathrm{H}}}{\mathrm{cm}^{-2}}) < 24$. The low number of counts and the high redshift of the source sample may lead to overestimate the intrinsic absorption. To account for this effect, we ran extensive X-ray spectral simulations and derived correction factors to apply to the observed distribution. Therefore, we obtained an estimate of the intrinsic N_{H} distribution. The results are consistent with a non-evolving obscured fraction of AGN with $L_{\mathrm{X}} \approx 10^{44}$ erg s^{-1} with redshift within 2σ.

(ii) The number counts of our $z > 3$ AGN sample are consistent with a decline in the AGN space density at high redshift, extending at faint soft fluxes ($F_{0.5-2\,\mathrm{keV}} \approx 4 \times 10^{-17}$ erg cm^{-2} s^{-1}) the behaviour determined at brighter fluxes found by Brusa et al. (2009) and Civano et al. (2011) in the COSMOS field. We used a more conservative approach with respect to Fiore et al. (2012) and Lehmer et al. (2012): if we significantly relax our selection criteria, our results are in a relatively good agreement at bright fluxes with their data, while at the faint end the discrepancy is a factor of $\sim$2.

These results would be improved by increasing the size and the reliability of the sample (e.g. gathering additional reliable redshifts via ultradeep optical–IR spectroscopic campaigns) and the

[8] We assigned the soft X-ray fluxes reported in Xue et al. (2011) to the sources not included in the high-redshift sample for the computation of all shaded areas.

quality of the X-ray spectra (e.g. increasing the *Chandra* exposure in this field). The accuracy of the photometric redshifts would be significantly improved by adding the deep Cosmic Assembly Near-IR Deep Legacy Survey (CANDELS) *H*- and *J*-band data (Grogin et al. 2011; Koekemoer et al. 2011).

ACKNOWLEDGMENTS

We acknowledge J. Aird for providing the LADE model data, M. Mignoli for his help with optical spectra, E. Vanzella for providing some spectroscopic redshifts and T. Dahlen for providing his catalogue of photometric redshifts. We thank G. Hasinger for the helpful suggestions which improved this work. We acknowledge financial support from the agreement ASI-INAF I/009/10/0 and the PRIN-INAF-2011. WNB and BL acknowledge financial support CXC grant SP1-12007A and NASA ADP grant NNX10AC99G. YQX thanks the Youth 1000 Plan (QingNianQianRen) program and the USTC startup funding (ZC9850290195). FEB acknowledges support from Basal-CATA (PFB-06/2007), CONICYT-Chile (FONDECYT 1101024) and Chandra X-ray Center grant SAO SP1-12007B. This research has made use of data obtained from the Chandra Data Archive and software provided by the Chandra X-ray Center (CXC) in the application packages CIAO.

REFERENCES

Aird J. et al., 2010, MNRAS, 401, 2531
Alexander D. M., Bauer F. E., Chapman S. C., Smail I., Blain A. W., Brandt W. N., Ivison R. J., 2005, ApJ, 632, 736
Alexander D. M., Swinbank A. M., Smail I., McDermid R., Nesvadba N. P. H., 2010, MNRAS, 402, 2211
Alvarez M. A., Wise J. H., Abel T., 2009, ApJ, 701, L133
Arnaud K. A., 1996, Astronomical Data Analysis Software and Systems V, 101, 17
Avni Y., 1976, ApJ, 210, 642
Ball W. H., Tout C. A., Żytkow A. N., Eldridge J. J., 2011, MNRAS, 414, 2751
Ballantyne D. R., Everett J. E., Murray N., 2006, ApJ, 639, 740
Begelman M. C., 2010, MNRAS, 402, 673
Bianchi S., Maiolino R., Risaliti G., 2012, Adv. Astron., 2012, Article ID 782030
Blackburn J. K., 1995, Astronomical Data Analysis Software and Systems IV, 77, 367
Brandt W. N., Alexander D. M., 2010, Proc. Natl. Acad. Sci., 107, 7184
Brusa M. et al., 2005, A&A, 432, 69
Brusa M. et al., 2009, ApJ, 693, 8
Brusa M. et al., 2010, ApJ, 716, 348
Cardamone C. N. et al., 2010, ApJS, 189, 270
Cash W., 1979, ApJ, 228, 939
Civano F. et al., 2011, ApJ, 741, 91
Comastri A. et al., 2011, A&A, 526, L9
Cowie L. L., Songaila A., Barger A. J., 1999, AJ, 118, 603
Daddi E. et al., 2007, ApJ, 670, 173
Dahlen T. et al., 2010, ApJ, 724, 425
Damen M., Labbé I., Franx M., van Dokkum P. G., Taylor E. N., Gawiser E. J., 2009, ApJ, 690, 937
Elbaz D. et al., 2011, A&A, 533, A119
Fan X. et al., 2006a, AJ, 131, 1203
Fan X. et al., 2006b, AJ, 132, 117
Feruglio C., Maiolino R., Piconcelli E., Menci N., Aussel H., Lamastra A., Fiore F., 2010, A&A, 518, L155
Fiore F. et al., 2012, A&A, 537, A16
Gilli R., Comastri A., Hasinger G., 2007, A&A, 463, 79 (GCH07)
Gilli R., Comastri A., Vignali C., Ranalli P., Iwasawa K., 2010, in AIP Conf. Proc. Vol. 1248, X-ray Astronomy 2009: Present Status, Multi-Wavelength Approach and Future Perspectives. Am. Inst. Phys., New York, p. 359
Gilli R. et al., 2011, Mem. Soc. Astrono. Ital. Suppl., 17, 85
Grogin N. A. et al., 2011, ApJS, 197, 35
Hasinger G., 2004, Nucl. Phys. B: Proc. Suppl., 132, 86
Hasinger G., 2008, A&A, 490, 905
Hasinger G., Miyaji T., Schmidt M., 2005, A&A, 441, 417
Hopkins P. F., Hernquist L., Cox T. J., Kereš D., 2008, ApJS, 175, 356
Iwasawa K. et al., 2012, A&A, 537, A86
Joye W. A., Mandel E., 2003, Astronomical Data Analysis Software and Systems XII, 295, 489
Kalberla P. M. W., Burton W. B., Hartmann D., Arnal E. M., Bajaja E., Morras R., Pöppel W. G. L., 2005, A&A, 440, 775
Koekemoer A. M. et al., 2011, ApJS, 197, 36
La Franca F. et al., 2005, ApJ, 635, 864
Lehmer B. D. et al., 2012, ApJ, 752, 46
Luo B. et al., 2008, ApJS, 179, 19
Luo B. et al., 2010, ApJS, 187, 560
Madau P., Rees M. J., 2001, ApJ, 551, L27
Maiolino R. et al., 2012, MNRAS, 425, L66
Mortlock D. J. et al., 2011, Nat, 474, 616
Norman C. et al., 2002, ApJ, 571, 218
Popesso P. et al., 2009, A&A, 494, 443
Rafferty D. A., Brandt W. N., Alexander D. M., Xue Y. Q., Bauer F. E., Lehmer B. D., Luo B., Papovich C., 2011, ApJ, 742, 3
Salvato M. et al., 2009, ApJ, 690, 1250
Salvato M. et al., 2011, ApJ, 742, 61
Santini P. et al., 2009, A&A, 504, 751
Silverman J. D. et al., 2010, ApJS, 191, 124
Spergel D. N. et al., 2003, ApJS, 148, 175
Steffen A. T., Barger A. J., Cowie L. L., Mushotzky R. F., Yang Y., 2003, ApJ, 596, L23
Sturm E. et al., 2011, ApJ, 733, L16
Szokoly G. P. et al., 2004, ApJS, 155, 271
Taylor E. N. et al., 2009, ApJS, 183, 295
Thomas D., Maraston C., Bender R., Mendes de Oliveira C., 2005, ApJ, 621, 673
Tozzi P. et al., 2006, A&A, 451, 457
Treister E., Urry C. M., 2006, ApJ, 652, L79
Trichas M. et al., 2012, ApJS, 200, 17
Turner T. J., George I. M., Nandra K., Mushotzky R. F., 1997, ApJS, 113, 23
Ueda Y., Akiyama M., Ohta K., Miyaji T., 2003, ApJ, 598, 886
Vanzella E. et al., 2008, A&A, 478, 83
Vanzella E. et al., 2010, ApJ, 725, 1011
Volonteri M., Begelman M. C., 2010, MNRAS, 409, 1022
Whalen D. J., Fryer C. L., 2012, ApJL, 756, L19
Willott C. J., McLure R. J., Jarvis M. J., 2003, ApJ, 587, L15
Willott C. J. et al., 2009, AJ, 137, 3541
Wuyts S., Labbé I., Schreiber N. M. F., Franx M., Rudnick G., Brammer G. B., van Dokkum P. G., 2008, ApJ, 682, 985
Wuyts S., van Dokkum P. G., Franx M., Förster Schreiber N. M., Illingworth G. D., Labbé I., Rudnick G., 2009, ApJ, 706, 885
Xue Y. Q. et al., 2011, ApJS, 195, 10
Yaqoob T., 1997, ApJ, 479, 184

APPENDIX A: DERIVING THE CORRECTION FACTORS FOR THE OBSERVED $N_{\rm H}$ DISTRIBUTION FROM SIMULATIONS OF X-RAY SPECTRA

The correction factors discussed in Section 3.3 are derived from spectral-fitting simulations using XSPEC, assuming five different degrees of obscuration (see Section 3.3 and Table 2). The five cases are simulated by runs of 1000 simulations each. The spectra are simulated at $z = 4$ with an input normalization tuned to obtain a distribution of net counts peaked at $\approx$100 (with a resulting dispersion of $\sigma \approx 11$ per cent), using the response files of a real source at an off-axis angle $\theta = 4$ arcmin. The results do not vary significantly if different response files are used, but they are sensitive to the number

of net counts and redshift. To be consistent with the real cases, for these parameters we assumed values close to the average ones of the sample.

The simulation procedure can be summarized in the following steps.

(i) Definition of an input model: we considered a power law with photon index fixed to $\Gamma = 1.8$ as input model, absorbed by five different column densities, whose values vary between the five considered cases (Table 2, third column). Galactic absorption ($N_{\rm H} = 7 \times 10^{19}\,{\rm cm}^{-2}$) is also included. The desired number of counts is obtained by properly adjusting the normalization of the power law.

(ii) Simulation of the spectrum: the *fakeit* command in XSPEC can simulate a spectrum with background, given a starting model. It requires as input the response and ancillary files, a real background file (with the POISSERR keyword set to TRUE) and the desired exposure time. We set it to 3.5×10^6 s to be consistent with the average real case. The output consists of a fake source and a fake background spectrum file. The same BACKSCAL keyword, read from the header of the input background file, is assigned by the *fakeit* procedure to both simulated spectra, which are therefore assumed to be extracted from regions with an equal area. This approach is not correct and causes the source and background counts to be wrongly scaled (i.e. a strong underestimate of the source net-count rate). To address this problem, before simulating the spectra we multiplied the EXPOSURE keyword of the real background file by a factor equal to the ratio between the BACKSCAL parameters of the real background and associated source file. We checked that the errors on the net-count rate were correctly computed (i.e. propagating the errors of the total and background-only count rates).

(iii) Spectral fitting: after having grouped the simulated spectra, using the FTOOL *grppha*, with at least 1 count per bin, and assigned to them the response files and the fake background spectra, the spectra are repeatedly fitted in the energy range $E = 0.5{-}5$ keV with an absorbed power-law model with $N_{\rm H}$ free to vary and photon index fixed to $\Gamma = 1.8$, accounting also for the Galactic absorption (fixed to $N_{\rm H} = 7 \times 10^{19}\,{\rm cm}^{-2}$). Cash statistics were employed in the fitting. The fitting step (which use a local minimization algorithm) is alternated to the computation of the errors of the various free parameters (which is a more global operation) for a few times, in order to avoid the fit to be stuck in a local minimum.

For each of the five considered cases of input $N_{\rm H}$, we obtained as output 1000 best-fitting values of $N_{\rm H}$. Following the procedure adopted to fit the real spectra, we define $N_{\rm H}$ to be constrained if the lower limit (at the 90 per cent CL) on the best-fitting value is larger than zero, otherwise we consider its upper limit (e.g. if the fit returns a best-fitting value $N_{\rm H} = 10^{+12}_{-10} \times 10^{22}\,{\rm cm}^{-2}$, we consider $N_{\rm H} < 22 \times 10^{22}\,{\rm cm}^{-2}$).

The probability p_{ij} that a simulated source in the jth $N_{\rm H}$ bin is observed in the ith $N_{\rm H}$ bin is derived by counting the number of times a constrained best-fitting $N_{\rm H}$ falls in bin i and normalized this number to 1000. In order to conserve the number of sources when the correction factors are applied on the real distribution, the p_{ij} factors must be rescaled to the total probability to constrain $N_{\rm H}$ (i.e. we excluded all cases in which only upper limits could be found):

$$P_{ij} = \frac{p_{ij}}{\sum_{i=A}^{E} p_{ij}}. \tag{A1}$$

As discussed in Section 3.3, bins A, B and C are merged into a single bin (ABC), since they are indistinguishable at $z > 3$. Therefore, the probability factors, $P_{i({\rm ABC})}$, are as follows:

$$P_{({\rm ABC})({\rm ABC})} = \sum_i \frac{\sum_j p_{ij}}{3} \tag{A2}$$

for $i, j = {\rm A, B, C}$; $P_{i({\rm ABC})} = \frac{p_{i{\rm A}} + p_{i{\rm B}} + p_{i{\rm C}}}{3}$; and $P_{({\rm ABC})j} = \frac{p_{{\rm A}j} + p_{{\rm B}j} + p_{{\rm C}j}}{3}$ for $i, j = {\rm D}$ and E.

The same procedure is then repeated using $\Gamma = 1.6$ instead of $\Gamma = 1.8$. The P_{ij} factors, for i and $j = ({\rm ABC})$, D and E, are those used in Section 3.3.

APPENDIX B: X-RAY SPECTRA OF THE $z > 3$ SAMPLE

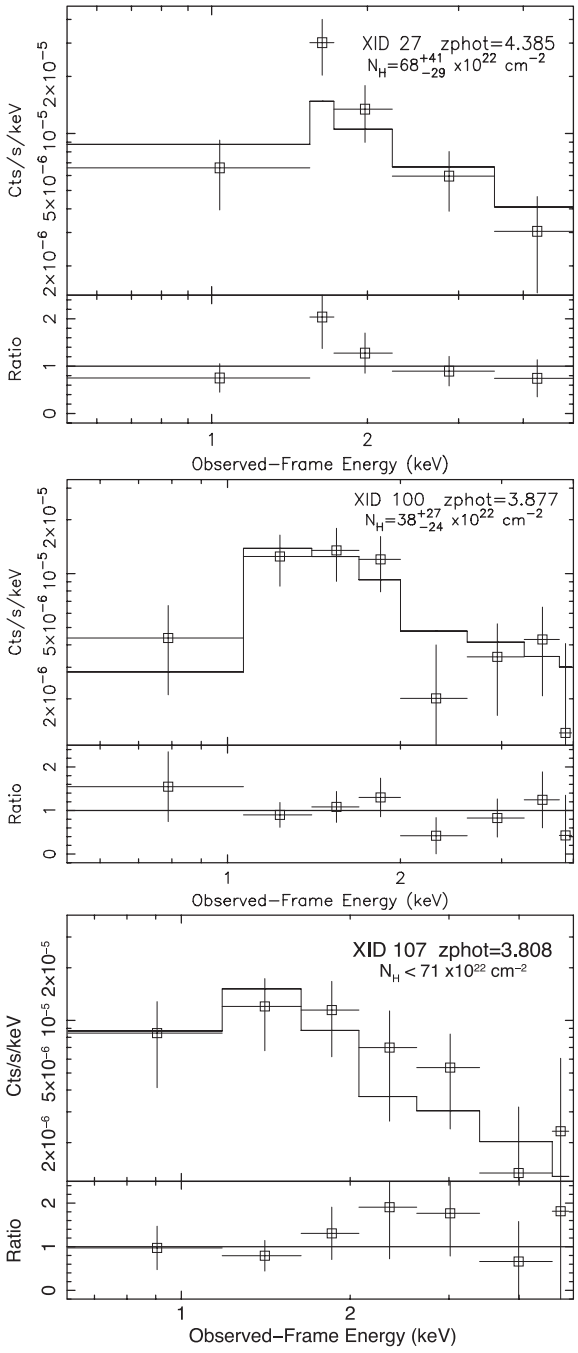

Figure B1. X-ray spectra of the 34 sources in the high-redshift sample, fitted with an absorbed power-law model with $\Gamma = 1.8$ (see Section 3.2). The ratio of data to folded model is shown and the spectrum is binned (at $\geq 3\sigma$ detection significance) for display purposes only. Identification numbers are from Xue et al. (2011). The spectroscopic or photometric redshift is reported for each source, as well as the best-fitting column density, accordingly to Table 1.

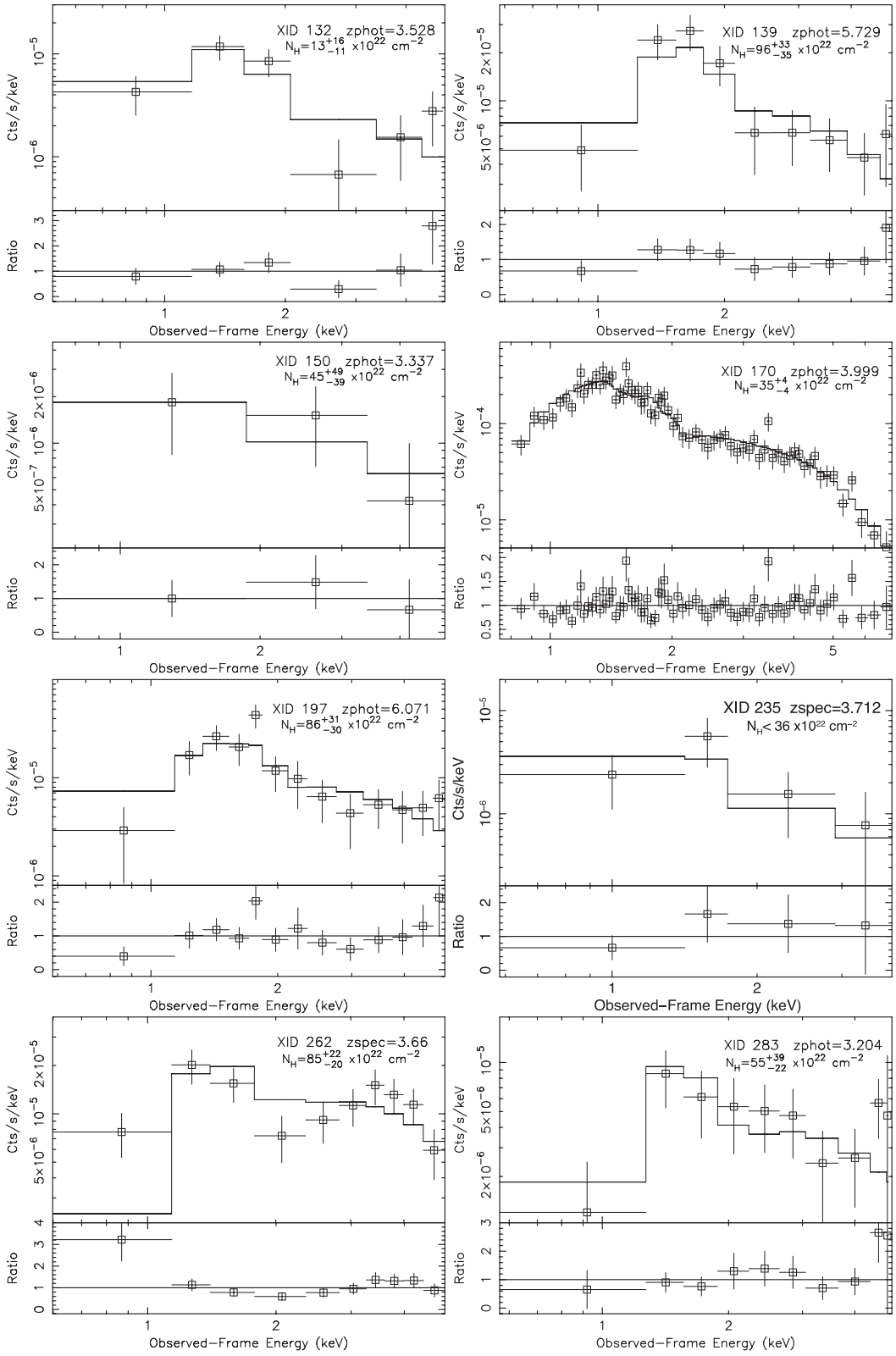

Figure B1 – *continued*

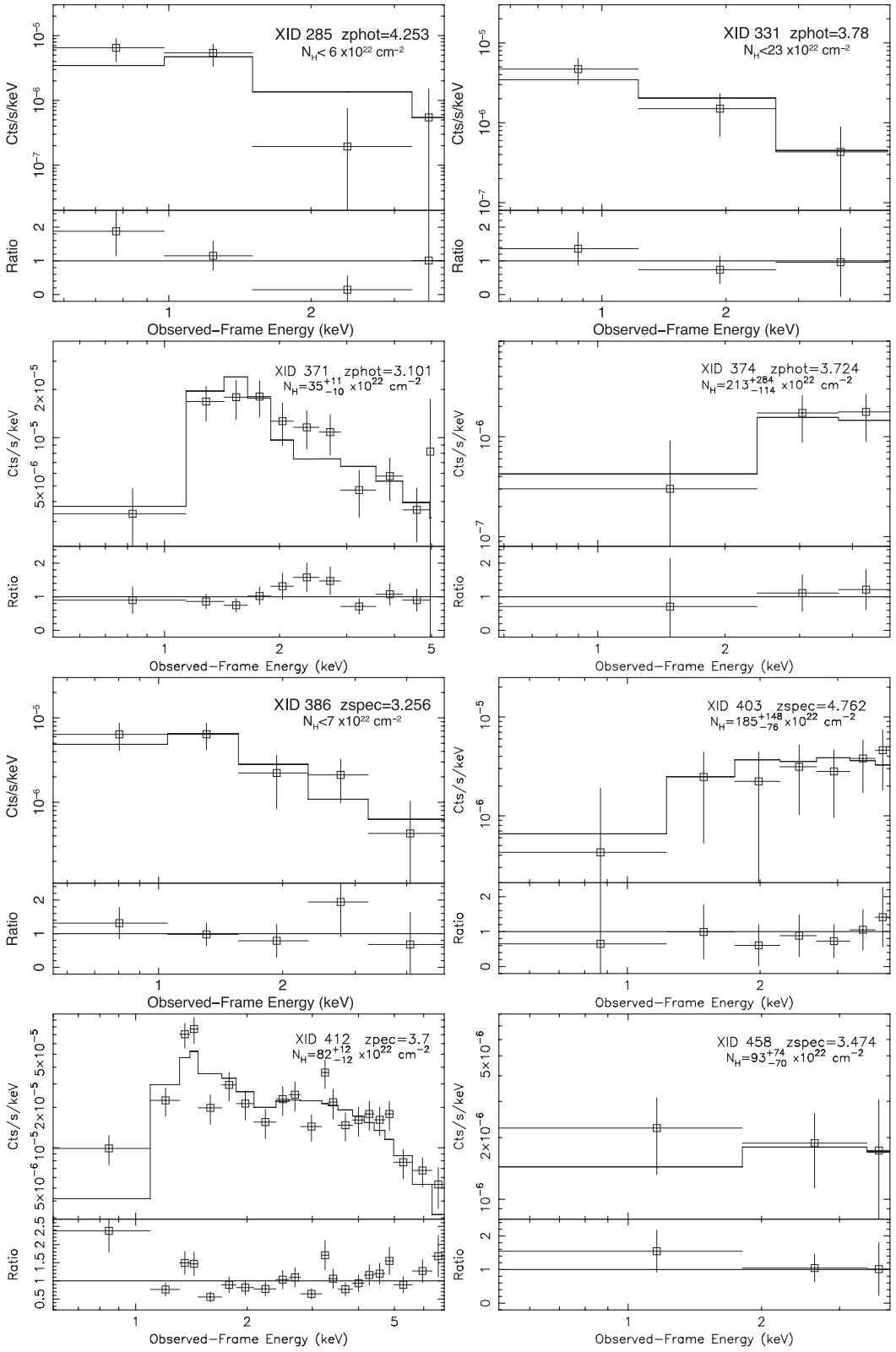

Figure B1 – *continued*

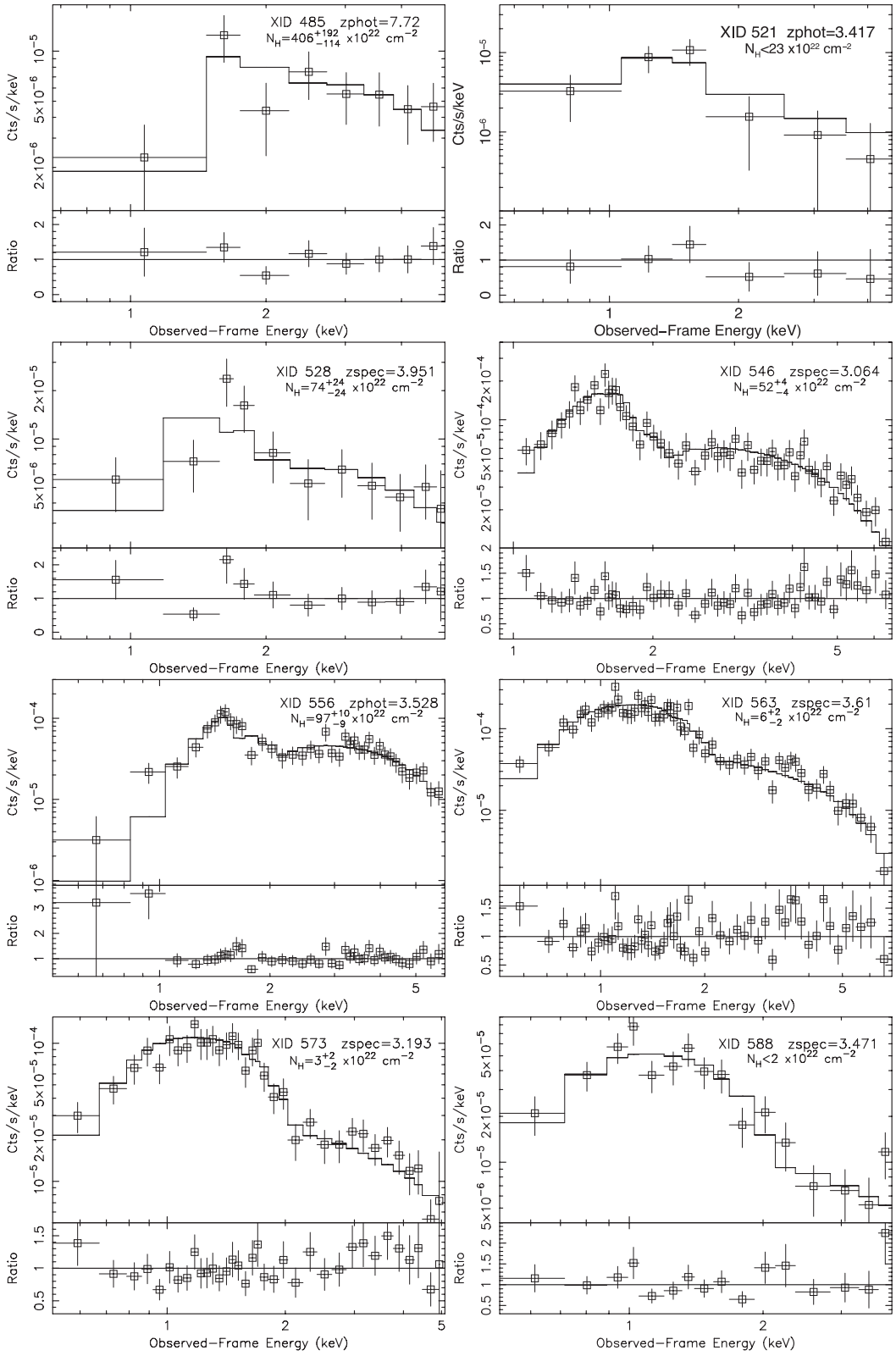

Figure B1 – *continued.*

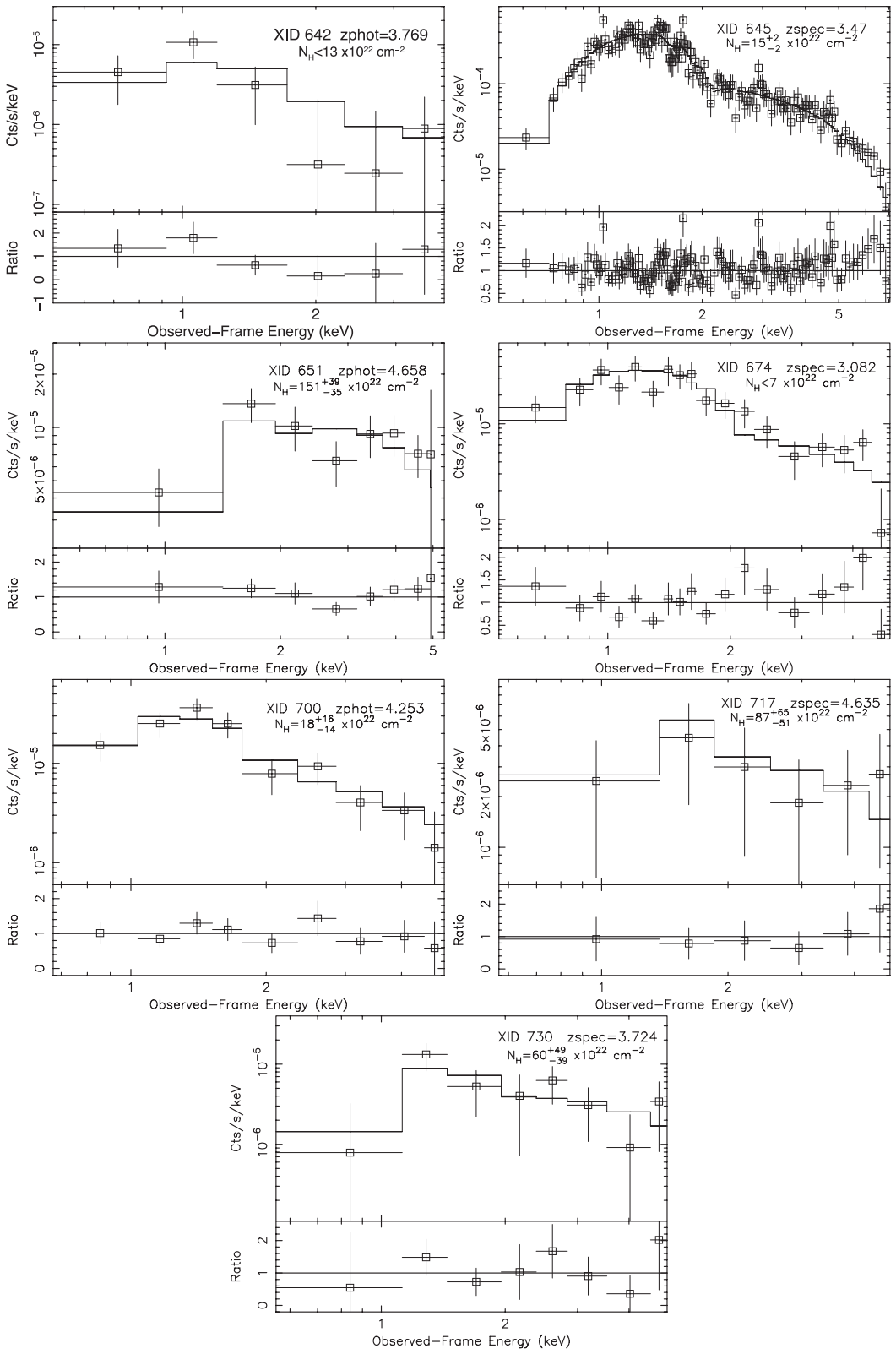

Figure B1 – *continued*

This paper has been typeset from a TEX/LATEX file prepared by the author.